Northern Ireland Assembly

Research and Information Service
Briefing Note

Paper 38/13 15 January 2013 NIAR 842-12

Robert Barry

Equality Issues in Recruitment to the Northern Ireland Civil Service (Update)

Summary of Briefing Paper (presented to Committee on 28 November 2012)

Analysis of recruitment competition outcomes in recent years suggests that some consistently occurring inequalities may exist within the Northern Ireland Civil Service (NICS) recruitment process.

Examination of equality monitoring information made available by the Department of Finance and Personnel, in relation to the results of selection tests used in recruitment competitions, suggests that females, Catholics, disabled and older candidates are, in general, less likely than their counterparts to make it through to the next stage of the competition.

While the Department has provided a fairly robust response in relation to its selection procedures and the actions it has taken to minimise the risk of bias, the question of whether or not the selection tests contain inherent bias remains unanswered.

Furthermore, it would appear that the adverse impact of NICS selection tests for females and Catholics (particularly in relation to the Executive Officer II competitions) has remained consistent and unchecked for at least 10 years. From the Department's response, it is not clear what specific action is going to be taken to investigate the reasons for this.

The available data also suggests that the Department might want to look more closely at the interview performance of different groups (where applicable), particularly with regard to gender, age and disability, to see if there is any consistent adverse impact occurring during that part of the selection process.

Further Information Provided by the Department in Relation to Selection Tests

The Department has since provided further information in relation to the selection tests used in recent competitions.

The tests cover a variety of skills (at different levels) and were supplied by a number of different suppliers (see Table 1 below). Some sample questions are available on SHL's web-site at www.shldirect.com/example_questions.html .

Table 1. Tests Used in Recent NICS Recruitment Competitions

Grade	Competition	Test Supplier	Test
AA and AO [n.b. same tests used for both competitions]	2008	Capita Resourcing	Numerical task
			Following Procedures
			Speed and Accuracy
EO2	2007	SHL	Verbal Evaluation from the Critical Reasoning Test Battery
		SHL	Data Interpretation from the Critical Reasoning Test Battery
EO2	2011	SHL	SHL online test - Verbal and Numerical Reasoning
		OPP	Critical Business Planning from the ABLE series [n.b. also used in the degree holder competitions]
		Psytech	Verbal Reasoning from the GRT2 General Reasoning test
		SHL	Data Interpretation from the Critical Reasoning Test Battery [also used in 2007]
Degree Holder	2007	OPP	Critical Business Planning from the ABLE series
		OPP	Critical Reasoning Skills Series (CRSS) – Verbal Reasoning
		NICS	NICS Financial Skills
Degree Holder	2009	NICS (answer sheets provided by SHL)	Financial Skills Test
		NICS	Verbal Test
		Pearson	Watson Glaser Critical Thinking Appraisal UK Version
		OPP	Critical Business Planning from the ABLE series
Graduate fast stream	2011, 2012, 2013	Parity (on behalf of Cabinet Office)	Initial shortlisting: Online numerical test Online verbal test Online competency questionnaire n.b. candidates who need adjustments because of their disabilities are not required to take either the numerical or verbal tests.

Source: Corporate Human Resources, Department of Finance and Personnel

Technical data was available only for the SHL on-line tests used in the 2011 EO II competition. SHL analysed their verbal and numerical ability test scores by sex, ethnicity and age using trials data obtained from 8,436 participants. From this analysis, they concluded the following:

> *"In summary, small to zero effect sizes were found for the verbal scores, while the effect sizes were obtained for the numerical scores indicate a small advantage for male candidates and for candidates 39 years or younger. A near zero effect size was obtained for the comparison of numerical scores by ethnicity."*

The effect sizes are presented in a standardised format (Cohen's *d*) as follows:

Table 2. Standardised effect sizes (*d* scores)

Test Type	Differences by sex	Differences by ethnicity	Differences by age
Verbal	0.06	0.11	0.04
Numerical	0.23	0.09	0.22

The figures in the table represent the differences between group means in terms of the number of standard deviations, where 0.3 and below is treated as small, 0.5 is treated as medium, and 0.8 and above is treated as a large effect size.

When the *d* values in the table above are converted into *t* values[1] (which take account of sample sizes), however, it is apparent that all the differences in the table except for one (verbal scores by age) are statistically significant, with probabilities as follows:

Table 3. Probabilities of obtaining these differences by chance (from *t* values)[2]

Test Type	Differences by sex	Differences by ethnicity	Differences by age
Verbal	$p < 0.01$	$p < 0.001$	$p > 0.05$
Numerical	$p < 0.001$	$p < 0.001$	$p < 0.001$

Percentile comparisons between the groups can also be calculated (from the *d* values in Table 2) as follows (showing the percentile positions for the average scores of males, whites and those aged 39 and under on the distribution of scores for females, non-whites and those aged 40 and older):

[1] Using $d = t(n1 + n2) / [\sqrt{(df)}\sqrt{(n1n2)}]$, where n1=4382 and n2=3885 for males and females; n1=3796 and n2=3796 for whites and non-whites; and n1=5155 and n2=3028 for the younger and older age groups respectively.

[2] For a two-tailed test.

Table 4. Percentile positions of average score for 'favoured' group on distribution of 'less favoured' group scores (from *d* scores)

Test Type	Differences by sex	Differences by ethnicity	Differences by age
Verbal	52	54	52
Numerical	59	54	59

What this broadly means is that if you took the median score for males as the pass mark on the numerical ability tests, 50 out of 100 males would pass and 41 out of 100 females would pass (i.e. a 22% higher pass rate for males).[3] On the verbal ability test, 50 out of 100 males would pass compared with 48 females out of 100 (i.e. a 4% higher pass rate for males). The outcomes for those aged 39 and under would be the same as for males (i.e. a 22% higher pass rate for the younger age group on the numerical ability test and a 4% higher pass rate on the verbal ability test). For whites, 50 out of 100 would pass on both tests compared with 46 out of 100 non-whites (i.e. a 9% higher pass rate for whites).

In conclusion, while the test effect sizes for this particular set of tests may be described technically as "small", it is difficult to see how they could be deemed small from a social equality perspective.

Both the Department and SHL have been open and helpful in providing this technical data and it may well be the case that the "small" bias inherent in these tests is no better or worse than that contained in similar tests produced by others in the market. It would be useful, however, if the technical data for group comparisons could be made available for other tests used by the Department for recruitment purposes.

Some analysis of the qualifications of applicants might also be useful to give an indication of any ability differences between different groups, particularly as it may be difficult to obtain technical data derived from test trials for religious differences.

Whatever the outcome of any further analysis, it is recommended that if tests are going to be used for selection purposes, then steps should be taken by test providers to neutralise test bias as far as possible, where statistically significant differences are found to exist.

[3] Assuming that the scores of both populations are normally distributed.